MvEcho - Acoustic Response Modelling for Auralisation

Léonie Buckley, Sam Caulfield, David Moloney
Movidius Ltd., 1st Floor, O'Connell Bridge House, D'Olier St, Dublin D02 RR99, Ireland
Email: {firstname.surname@movidius.com}

Introduction

- Auralisation is the process of simulating the listening experience at a given position.
- While many auralisation algorithms currently exist, the majority require a manual description of the environment.
- In contrast, MvEcho removes the need for this manual description to provide an environment independent algorithm to approximate auralisation.
- Initial implementation of MvEcho concerned the acoustic response of objects, namely the pyramid El Castillo, (shown below in Fig. 1) and has now been extended to model the response of rooms.
- The algorithm is fully functional in C and is currently being ported to the Myriad 2.

VOLA

- Voxels are used to represent regular rectangular grids in 3D space, the 3D equivalent to pixels.
- Voxel models are implemented using VOLA (Volumetric Accelerator), a software library developed by Movidius[1].
- VOLA is extremely memory efficient – storing either a single or two bits per occupied voxel.
- The resolution of the environment is specified by the user.
- Models can be imported into VOLA from Blender models.
- Technologies such as the Microsoft Kinect are available to scan, detect and measure objects in 3D space. Such scans can be used to obtain a voxel representation using binvox and VOLA.

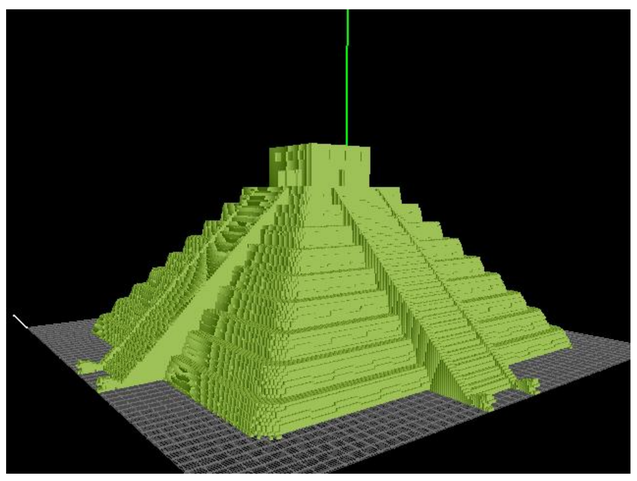

Fig. 1: El Castillo

Use Cases

- Modelling of audio for VR – The user's surrounding are often neglected when providing audio for VR. MvEcho could model the acoustic response of the environment to allow for more realistic VR audio.

Fig. 2: VR headset

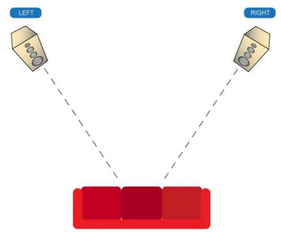

Fig. 3: Speaker calibration

- Calibration of speakers and soundbars based on the environment and listener and speaker locations.

Ray Casting

- A single ray is cast out from the sound source towards each visible occupied voxel, and its path is then recorded back to the observer's head. These rays emulate early reflections.
- Late reverberations are not yet taken into account, but Fig. 4 below shows that the majority of the sound energy is carried in direct sound and early reflections.

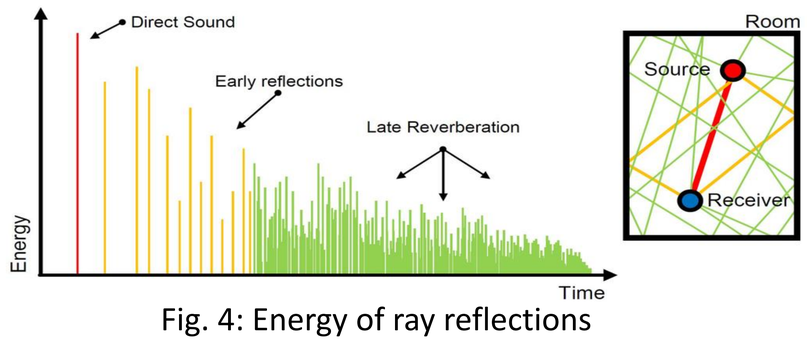

Fig. 4: Energy of ray reflections

- The response of the environment is modelled using the information obtained from the ray casting.
- The input stimulus is an approximation of an impulse such as a handclap. The user can specify the stimulus and observer locations.
- The attenuation and delay of each sample of the stimulus is calculated.
- MvEcho acts as an FIR filter to combine the response of each sample to emulate the acoustic response of the environment, as shown below in Fig. 5.

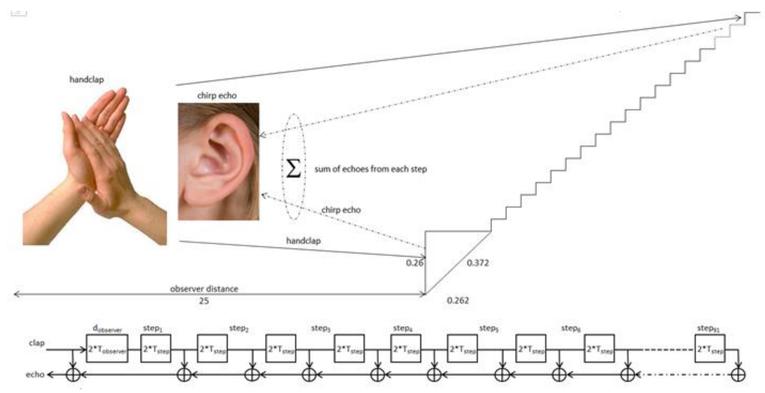

Fig. 5: Listener and sound source locations

Myriad 2

- The Myriad 2 MA2150 VPU is a heterogeneous, multicore always-on System-on-Chip.
- Myriad2 is based on the proprietary 128-bit SHAVE vector-processor (x12) and hardware acceleration pipeline backed by shared multicore memory subsystem and peripherals.
- It occupies 27mm^2 in 28nm HPM-CMOS and has been designed to operate at 0.9V for nominal 600MHz operation.
- Porting of MvEcho to the Myriad2, shown below in Fig. 6, is in progress.
- The Myriad 2's 12 SHAVE processors provide parallel processing, which allows the acoustic response of multiple slices of the environment to be computed in parallel.

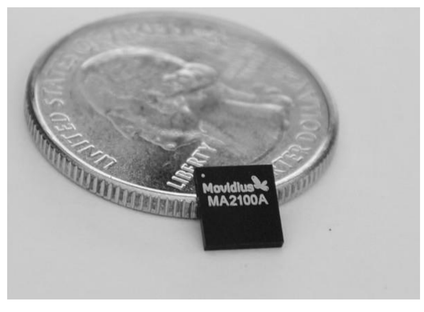

Fig.6: Myriad 2

References

1. Xu et al (2016) '3D Object Recognition Based on Volumetric Representation Using Convolutional Neural Networks', in Articulated Motion and Deformable Objects. Springer International Publishing Switzerland